DOCUMENT RESUME

ED 330 707 TM 016 275

AUTHOR Edwards, Lynne K.
TITLE Applying the Generalized Sequential Bonferroni Method to a Set of Planned and Post Hoc Comparisons.
SPONS AGENCY Minnesota Supercomputer Inst.
PUB DATE Apr 91
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Equations (Mathematics); *Mathematical Models; Statistical Analysis
IDENTIFIERS *Bonferroni Procedure; *Planned Comparisons; Post Hoc Methods; Weighted Data

ABSTRACT

This study analytically examined the multiple comparison procedures for testing a combined set of planned and post hoc comparisons. To establish a practical guideline, the following issues were investigated: (1) when the generalized sequentially rejective Bonferroni method (GSRB) is and is not applicable; (2) the required alpha levels for the GSRB compared with six alternatives (classical Bonferroni (CB)-CB, CB, sequentially rejective Bonferroni (SRB)-SRB, SRB, planned comparison-SRB, and a procedure proposed by R. Rosenthal and D. B. Rubin (1984); and (3) the optimal weighting system for the GSRB. When a combined set of planned and post hoc comparisons is tested in the same study, the GSRB can be used with W, which represents the weight for each planned comparison set to n, the number of post hoc comparisons (i.e., W=n if n is greater than or equal to 2, otherwise W=2 and W=1 for the post hoc comparisons). If the planned comparisons are orthogonal and each is tested at the per comparison level, the GSRB is not recommended. Instead, the SRB can be applied to the post hoc portion of the comparisons. A numerical example with data for four groups was used to illustrate seven multiple comparison procedures. Although the procedures of controlling for the inflated Type I error arising from the multiplicity of tests remain controversial, the GSRB seems promising in conducting a combined set of planned and post hoc comparisons. Four tables contain illustrative data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LYNNE K. EDWARDS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED330702

Applying the generalized sequential Bonferroni method to a set of planned and post hoc comparisons

Paper presented at the annual meeting of the American Educational Research Association in Chicago (April , 1991).

Lynne K. Edwards

Department of Educational Psychology

University of Minnesota

Running Head: Generalized Bonferroni

The author is grateful to the Minnesota Supercomputer Institute for the funds. Requests for reprints should be sent to Lynne K. Edwards, Dept. of Educational Psychology, University of Minnesota, 323 Burton Hall, 178 Pillsbury Dr SE, Minneapolis, MN 55455.

BEST COPY AVAILABLE

TM016275

Applying the generalized sequential Bonferroni method to a set of planned and post hoc comparisons

Abstract

When a combined set of planned and post hoc comparisons are tested in the same study, the generalized sequentially rejective Bonferroni method (GSRB) can be used with W, the weight for each planned comparison set to n, the number of post hoc comparisons, i.e., $W = n$, if $n \geq 2$ else $W = 2$, and $W = 1$ for the post hoc comparisons. If the planned comparisons are orthogonal and each is tested at the per comparison level, the GSRB is not recommended. Instead, the sequentially rejective Bonferroni (SRB) can be applied to the post hoc portion of the comparisons. A numerical example was used to illustrate seven multiple comparison procedures. Although the procedures of controlling for the inflated type I error arising from the multiplicity of tests remain controversial, the GSRB seems promising in conducting a combined set of planned and post hoc comparisons.

Key words: Bonferroni inequality, Sequentially rejective methods, Weighted procedures, post-hoc comparisons, planned comparisons, Multiple comparisons,

Introduction

Often there are occasions in which both planned and post hoc multiple comparisons are made within the same study. For example, a subset of the experimental conditions can be a replication from previous studies while some conditions may be added to be explored. Unfortunately, there has not been a clear statistical guideline for dealing with such a situation. Kirk (1982) recommends handling the planned portion and the unplanned portion separately, but the relative merits among the applicable procedures are not well known. Obviously, the nature of the hypotheses: pairwise or complex linear contrasts, and orthogonal or nonorthogonal, does influence which procedures are appropriate, but the classical Bonferroni (CB) and other Bonferroni type procedures have been proven to be useful in many different situations: equal or unequal sample sizes; pairwise or complex linear contrasts; all possible pairwise or selected number of comparisons. Furthermore, several modified Bonferroni procedures which are more powerful than the CB have been recently developed.

It may be argued that all the Bonferroni type procedures are applicable to the planned comparisons only. The Bonferroni inequality, however, still holds among the post hoc comparisons, although a caution is warranted in confirming the post hoc comparisons generated and tested within the same study (Saville, 1990). It is less known that the CB is recommended for testing the post hoc comparisons provided that there are a limited number of comparisons (Edwards, 1985); and for setting the simultaneous confidence intervals for a combined set of planned and post hoc comparisons (Ury & Wiggins, 1971, 1974; Rodger, 1973). In the literature on multiple comparisons, all possible pairwise comparisons are often used in examining the power of the Bonferroni type procedures (Shaffer, 1986; Holland & Copenhaver, 1987; Seaman, Franke, Serlin, & Levin, 1989; Kim & Edwards, 1990), which are hardly defendable as "planned" or "guided by theory."

There are two types of Bonferroni type procedures: One type is the simultaneous procedures such as the one proposed by Rosenthal and Rubin (1984), RR, the CB and the slight improvement of the CB (Simes, 1986; Worsley, 1982); and another type is the

sequential procedures such as the two proposed by Holm (1979), the SRB and GSRB, and the two modified sequential procedures (Shaffer; 1986; Holland & Copenhaver; 1987).

The relative power among these Bonferroni type procedures is somewhat known. Among the CB, SRB, and GSRB, all of which control the experimentwise error rate (EW) to at most α, the order of power is GSRB $\geq$ SRB $\geq$ CB provided that there are no overlapping hypotheses (Holm, 1979). The GSRB is also shown to be more powerful than the RR (Holland & Copenhaver, 1988). Although there are other modified Bonferroni type procedures which are more powerful than the CB, RR, SRB, or GSRB, but they are practically limited to all possible pairwise comparisons in order to make use of the "logical implication" or "orthant dependency," neither of which is easily determined in a mixture of pairwise and complex linear contrasts nor is it easily modified for testing only a selected set of contrasts.

The procedures of controlling for the inflated type I error arising from the multiplicity of tests remain controversial, in particular, where differential "importance" weights are allocated among the planned comparisons (Ryan, 1985; Rosenthal & Rubin, 1984). Neither the GSRB or RR, however, has been applied to a situation where a combined set of planned and post hoc comparisons are made. If the two weights, one for the planned and another for post hoc comparisons, can be made objectively without "looking over the data," the weighted procedures can be potentially useful.

Seven practical procedures with general applications, the CB, CB-CB, SRB, SRB-SRB, PC-SRB, GSRB, and RR, are examined in this study. The CB, SRB, GSRB, and RR methods apply one system to the whole set of planned and post hoc comparisons, while the two-stage procedures such as the CB-CB, SRB-SRB, and PC-SRB apply separate systems for the planned portion and for the post hoc portion. The notations used are: $c = m + n$, where m is the number of planned comparisons, n is the number of post hoc comparisons, and c is the total number of comparisons; the ith hypothesis will be examined at $\alpha_{(i)}$ level and the experimental error rate (EW) is set to α; and $W(i)$ is the weight for the ith hypothesis. The α level applied to each of the hypotheses by these seven procedures are examined as follows:

CB: None of the hypotheses are sorted by their p values. In essence, the more hypotheses are tested, the equally strict α level is assigned to both planned and post hoc hypotheses. Each hypothesis, regardless of whether it is planned or post hoc will be examined at $\alpha_{(i)} = \alpha / c$ for i = 1, ..., c.

CB-CB: None of the hypotheses are sorted by their p values. Typically more post hoc hypotheses will be examined and they will be tested at a smaller α level than that of the planned hypotheses. Each hypothesis will be examined at the individual α level so that two separate familywise α's (one for planned and another for post hoc) will not exceed the preset experimentwise error rate.
$\alpha_{(i)} = (\alpha / 2) / m$ for i = 1, ..., m; and $\alpha_{(i)} = (\alpha / 2) / n$ for i = 1, ..., n.

SRB: All the hypotheses combined are sorted by their p values. In this system, a large planned contrast may be tested at a smaller α level than a small post hoc contrast. Each hypothesis, regardless of whether it is planned or post hoc will be examined at
$\alpha_{(i)} = \alpha / \sum_{j=i}^{c} 1$ for i = 1, ..., c.

SRB-SRB: The planned and post hoc hypotheses are separately sorted by their p values. In this system, the potential bias for post hoc contrasts found in the SRB system is eliminated. Aside from the PC-SRB, the SRB-SRB is most powerful for the planned comparisons among the equally weighted procedures considered. Each hypothesis will be then examined at
$\alpha_{(i)} = (\alpha/2) / \sum_{j=i}^{m} 1$ for i = 1, ..., m; and $\alpha_{(i)} = (\alpha/2) / \sum_{j=i}^{n} 1$ for i = 1, ..., n.

PC-SRB: The post hoc hypotheses only are sorted by their p values. This procedure gives a larger α allocation to the planned hypotheses, while apply the SRB to only the post hoc portion. Each hypothesis, depending on whether it is planned or post hoc will be examined at $\alpha_{(i)} = \alpha / 2$ for i = 1, ..., m; and $\alpha_{(i)} = (\alpha / 2) / n$ for i = 1, ..., n.

RR: The hypotheses combined are not sorted by their p values. In this system, the determination and assignment of the proper weight, *W*(i), is the most difficult portion of the

procedure. Each hypothesis, regardless of whether it is planned or post hoc will be examined at

$$\alpha_{(i)} = \alpha W_{(i)} / \sum_{j=1}^{c} W_{(j)} \quad \text{for } i = 1, ..., c.$$

GSRB: All the hypotheses combined are sorted by their p/W values. By assigning $W = 1$ for each of the post hoc comparisons, W for the planned hypotheses can be solved by equating the power of this test to some other multiple comparison method. The planned comparisons portion can be set to as powerful as the SRB-SRB by assigning the α level for the ith planned comparison for the GSRB equal to or larger than that of the SRB-SRB. That is, $\alpha_{(i)} = \alpha/2(m - i + 1) \leq \alpha W_{(i)} / n + (m - i + 1)W_{(i)}$. By solving the inequality, $W \geq n$ for $n \geq 2$, else $W \geq 2$ for the planned comparisons is obtained. Each hypothesis, regardless of whether it is planned or post hoc will be examined at

$$\alpha_{(i)} = \alpha W_{(i)} / \sum_{j=i}^{c} W_{(j)} \quad \text{for } i = 1, ..., c.$$

Purpose of the Study

The purpose of this study is to analytically examine the multiple comparison procedures for testing a combined set of planned and post hoc comparisons. To establish a practical guideline, this study investigates (1) when the GSRB is applicable and when it is not; (2) the required α levels for the GSRB in comparison to six alternatives: CB-CB, CB, SRB-SRB, SRB, PC-SRB, and RR; and (3) the optimal weighting system for the GSRB.

Method

The required probability levels among the seven multiple comparison procedures are analytically examined by setting EW to be at most α. Although it is possible to unequally allocate EW between the planned and post hoc portions, it is rarely justifiable without any prior information. Therefore, EW is equally allocated between these two portions in this study. The optimal weights for the GSRB is derived algebraically. Because all procedures result in t-values with the same degrees of freedom, it is sufficient to examine the required probability

levels among the procedures to determine the relative test sizes and power. A numerical example (Kirk, 1982) is used to illustrate the applications of these seven multiple comparison procedures.

Analytical results

Selected results are summarized in Table 1. By choosing the common weight for the planned comparisons equal to the number of post hoc comparisons, W = n if $n \geq 2$, else W = 2, and W = 1 for the post hoc comparisons, the GSRB provides at least the same power for the planned hypotheses as the SRB-SRB. Although it is possible to adopt W > n, it reduces the power for the post hoc contrasts. The benefits of applying the GRSB are: (1) it is flexible and versatile for pairwise as well as complex linear contrasts ; (2) it is a computationally simple one-stage method compared to the two-stage methods (e.g., SRB-SRB); (3) it assures the same or better power level for the planned comparisons as the SRB-SRB without excessively reducing the average detection power for the post hoc comparisons;and (4) it keeps EW to at most α for the set of all hypotheses. If the planned comparisons are orthogonal and are tested at the per comparison level (i.e., $\alpha/2$ each), then there is no weighting system for the GSRB which will provide the equivalent power for the GSRB. The SRB can be applied to the post hoc portion of the comparisons instead.

Insert Table 1 about here

A numerical example

These seven procedures are illustrated using the data presented in Kirk (1982, p. 141) on a completely randomized group design with four levels of Treatment. Kirk illustrated an example where one hypothesis: Γ: $\mu_3 - \mu_4 = 0$, was planned, thus tested at some assigned a level. Subsequently, the overall F test was conducted on the residual SStreat with the corresponding df before the post hoc comparisons were made. This illustration may be plausible if a set of post hoc comparisons and a set of planned comparisons are orthogonal to each other. If any of

the post hoc hypotheses is nonorthogonal to any one of the planned hypotheses, the overall F test on the residual, i.e., SStreat - SSplanned, cannot assure consistency in the multiple comparison method even if a consistent method such as Scheffe's test is used. Such an example can be found where a subset of pairwise tests are planned and the rest pairwise tests are unplanned. A similar difficulty can arise, for example, in applying Tukey's test to the post hoc portion only. Tukey's largest range test has to be applied to all groups even if a subset of the pairwise tests are conducted as post hoc. These problems give added support for choosing multiple comparisons that incorporate both planned and post hoc portions.

To illustrate how to apply these seven procedures to a data set, an example data with four groups were analyzed (Kirk, 1982, p. 141). The summary table and the groups means were presented in Tables 2 and 3 (Kirk, pp. 141 & 148). We assume that two pairwise contrasts were planned (Groups 3 and 4; and Groups 2 and 4) and the rest of the pairwise contrasts were post hoc. Using the CONTRAST command in the ONEWAY SPSSx procedure, each contrast was examined and its significance level was reported assuming each was tested at the PC level.

Insert Tables 2 and 3 about here

Unfortunately, the computer printout does not report any significance level smaller than 0.001, which makes it slightly difficult to obtain more precise results. In this example, however, all comparisons except for the one between Groups 1 and 2 had sizable differences, which in turn obscured the differences among these seven procedures. The results summarized in Table 4 show the t value for each contrast as well as the actual ordering found among six comparisons. Comparing the results in Table 4 along with the nominal level of α for each individual contrast listed in Table 1, the relative power potentials among these seven methods can be evaluated.

Insert T

There are occasions when the procedu
also applicable or more powerful. Howev
pairwise contrasts whereas the CB, SRB,
number of contrasts is known and that the
can be objectively determined by assignir
procedure for the set of planned comparis
planned comparisons are orthogonal and
should be cautious about examining the p
same study, the GSRB is a computational
clear, and flexible procedure that is applic
EW is kept to at most α in testing a comb
Although the procedures of controlling for
of tests remain controversial, the GSRB s
planned and post hoc comparisons.

Insert Table 4 about here

Discussion

procedures other than the ones examined in this study are
However, these alternatives are often limited to all possible
, SRB, and GSRB are more flexible provided that the total
that they are not overlapping. The weights for the GSRB
assigning at least the same power as the SRB-SRB
omparisons. The use of the GSRB is discouraged when the
al and are tested at the PC level. Although researchers
ng the post hoc hypotheses derived and tested within the
tationally simple, intuitive, reasonably powerful, logically
s applicable to both equal and unequal sample sizes in which
a combined set of planned and post hoc comparisons.
ling for the inflated type I error arising from the multiplicity
GSRB seems promising in conducting a combined set of

References

Edwards, A. L. (1985). Experimental design in psychological research (5th ed.). New York: Harper & Row.

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni test procedure. Biometrics, 43, 417-423.

Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. Psychological Bulletin, 104, 145-149.

Holm, S. (1979). Sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65-70.

Kim, G., & Edwards, L. K. (1990). Multiple comparisons of means with an emphasis on the recent developments. ED 317 597 Washington, D.C.: Dept of Education.

Kirk, R. E. (1982). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Wadsworth.

Rodger, R. S. (1973). Confidence intervals for multiple comparisons and the misuse of the Bonferroni inequality. British Journal of Mathematical and Statistical Psychology, 26, 58-60.

Rosenthal, R., & Rubin, D. B. (1984). Multiple contrasts and ordered Bonferroni procedures. Journal of Educational Psychology, 76, 1025-1034.

Ryan, T. A. (1985). "Ensemble-adjusted p values": How are they to be weighted? Psychological Bulletin, 97, 521-526.

Saville, D. J. (1990). Multiple comparison procedures: The practical solution. The American Statistician, 44, 174-180.

Seaman, M.A., Franke, T. M., Serlin, R. C., & Levin, J. R. (1989). New, improved multiple comparisons procedures: More pep with each step. Paper presented at the annual meeting of the American Educational Research Association (San Francisco, March, 1989).

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. Journal of the American Statistical Association, 81, 826-831.

Simes, R. J (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrics, 73, 751-754.

Ury, H. K., & Wiggins, A. D. (1974). Use of the Bonferroni inequality for multiple comparisons among means with post hoc contrasts. British Journal of Mathematical and Statistical Psychology, 27, 176-178.

Worsley, K. J. (1982). An improved Bonferroni equality and applications. Biometrika, 69, 297-302.

Table 1. A comparison of the required p-value for each contrast

	SRB-SRB*	CB-CB*	CB**	SRB**	GSRB** w(i)	RR** (a)	RR** (b)	PC-SRB
Planned								
ψ_1	.0125	.0125	.0082	.0082	4	.0167	.0250	.0167
ψ_2	.0250	.0125	.0082	.0100	4	.0250	.0500	.0167
Post hoc								
ψ_3	.0062	.0062	.0082	.0125	1	.0125	.0042	.0042
ψ_4	.0083	.0062	.0082	.0167	1	.0167	.0045	.0042
ψ_5	.0125	.0062	.0082	.0250	1	.0250	.0050	.0042
ψ_6	.0250	.0062	.0082	.0050	1	.0500	.0056	.0042

** $\alpha = 0.05$ for the combined planned and post hoc comparisons

SRB-SRB: p_1 (Probability for ψ_1) $\leq p_2$ and $p_3 \leq p_4 \leq p_5 \leq p_6$ is assumed.

SRB: $p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5 \leq p_6$ is assumed.

Case (a): $p_1/W_1 \leq p_2/W_2 \leq p_3/W_3 \leq p_4/W_4 \leq p_5/W_5 \leq p_6/W_6$ is assumed.

Case (b): $p_3/W_3 \leq p_4/W_4 \leq p_5/W_5 \leq p_6/W_6 \leq p_1/W_1 \leq p_2/W_2$ is assumed.

Table 2. ANOVA summary table for Type CR-4 Design (Kirk, 1982, p.141)

Source	*SS*	*df*	*MS*	*F*
Between	194.500	3	64.833	44.28
Within	41.000	28	1.464	

Table 3. Absolute values of differences among means (Kirk, 1982, p.148)

	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$	$\overline{Y}_{.3}$	$\overline{Y}_{.4}$
$\overline{Y}_{.1} = 2.75$	--	0.75	3.50	6.25
$\overline{Y}_{.2} = 3.50$		--	2.75	5.50
$\overline{Y}_{.3} = 6.25$			--	2.75
$\overline{Y}_{.4} = 9.00$				

Table 4. A numerical example for the data in Table 2

		p	SRB-SRB	CB-CB	CB	SRB	w(i)	GSRB	RR	PC-SRB
Planned										
$\psi 1$	0.001		*	*	*	*	4	*	*	*
$\psi 2$	0.001		*	*	*	*	4	*	*	*
Post hoc										
$\psi 3$	0.001		*	*	*	*	1	*	*	*
$\psi 4$	0.001		*	*	*	*	1	*	*	*
$\psi 5$	0.001		*	*	*	*	1	*	*	*
$\psi 6$	0.225		n.s.	n.s.	n.s.	n.s.	1	n.s.	n.s.	n.s.

$\psi 1 = \mu 2 - \mu 4$ t = -9.090
$\psi 2 = \mu 3 - \mu 4$ t = -4.545
$\psi 3 = \mu 1 - \mu 4$ t = -10.330
$\psi 4 = \mu 1 - \mu 3$ t = -5.785
$\psi 5 = \mu 2 - \mu 3$ t = -4.545
$\psi 6 = \mu 1 - \mu 2$ t = -1.240

For SRB and RR: The inequality $p3 \leq p1 \leq p4 \leq p2 \leq p5 \leq p6$ holds.
Note: EWα = 0.05 is used for the combined planned and post hoc comparisons. If significant, an asterisk is placed in its place.